CSC2730 Group Project
Due Dec. 8th by end of day
(Late submission will not be graded.)

(You may complete this project as a group of 2 persons.)

Identify a practical problem where model prediction can be applied. Try 1 or 2 models (you may try models from scikit-learn that are not discussed in the class), and investigate which model configuration/choice of hyper-parameters gives the best prediction for your problem.

Write a report on your investigation. In the report, clearly describe:
1. The problem you considered, including the data used, the attributes/features that are associated with each instance and the target of your prediction.
2. The models and the hyper-parameters that you tried.
3. Make a table to summarize the performance of the models and different choice of hyper-parameters. Use a performance measure such as accuracy or f1 score. State how the performance is measured, e.g., using 5-fold cross validation.
4. Give a short discussion on why the attributes and model(s) you choose give good/bad performance and what improvement can be done to make it better.

Project Submission
Submit your project in moodle. Your submission should include the following:
1. the project report as a PDF file (***Don't forget to put the names of the persons in your group on the report.***)
2. the code (one or more .py files)
3. a readme file describing how to run your code
4. some example data (Do not include all the data you used. Just a few examples.)
Zip the above files into a zip archive and submit it by uploading the zip file in moodle.

Each group needs to make only one submission. (We will know the group from the names on your report.)

Grading Criteria
1. Your report should give a clear description of the problem, the data, the attributes, the model(s), the performance result and have some discussion on the models' performance.
2. We should be able to run your code on the small example dataset.
3. Originality and creativity of the problem you considered and the prediction approach you use. (Simply using datasets from scikit-learn and then applying some models is not very creative.)